

# **Exploring Gender Gap and School Differential Effects in Mathematics in Chilean**

## **Primary Schools**

**Word count: 6,349**

### **Abstract**

Despite the systematic recognition of large gender gaps in favour of boys in mathematics performance in Chile, the role of schools in explaining this gap has been comparatively underresearched. This paper analyses a longitudinal dataset to explore within and between school variation of Chilean primary schools in promoting attainment and progress of girls and boys. In order to reflect the hierarchical nature of the educational data, Multilevel modelling is used, with 163,044 students, nested within 3,355 schools, within 310 Municipalities to fit Raw, Contextualised Attainment and Value-Added models. The weak evidence indicative of gender differential school effect in progress in Chilean primary schools was interpreted as non-substantive. However, differences in effectiveness played a significant role, as girls progressed more than boys in less effective schools, but the opposite was true in more effective schools. Finally, the study concludes that the Chilean gender gap needs to be addressed mainly beyond schools.

### **Keywords**

Differential effects, gender gaps, multilevel modelling, Value added, Chilean primary schools.

## **Introduction**

From a Human Capital perspective, gender equality is concerned with “ensuring that all learners, boys and girls irrespective of their sex, have access to the same resources, educational opportunities and education outcomes” (Aikman & Unterhalter, 2013, p. 51). In other words, education quality cannot be accomplished without gender equity, and the gender gap is a key indicator to monitor in order to promote greater equity in educational achievement. In this context, the analyses of large-scale international educational comparative studies –such as The Programme for International Student Assessment (PISA) from the Organisation for Economic Co-operation and Development (OECD) and Trends in International Mathematics and Science Study (TIMSS)- have consistently shown that gender gaps in mathematics achievement are not universal, but specific to the situations of girls and boys in particular countries. Else-Quest, Hyde and Linn (2010) have argued that over the last three decades “the gender gap in math persists in some nations but not in others” (p. 122). Importantly González de San Román and De La Rica (2012) also point out that “girls perform relatively better in both maths and reading in societies where gender equality is enhanced, and the effect varies over the distribution of scores” (p. 1).

The gender gap in mathematics has also been described as “an important and extremely divisive issue of academic debate” (Fryer & Levitt, 2010, p. 210), as theories trying to make sense of it dissent. Those supporting the “nature hypothesis” have stressed irreconcilable biological differences between girls and boys, while researchers supporting the “nurture hypothesis” have pointed to cultural, social and societal factors (such as the role of the media perpetuating or transforming stereotypes about gender and mathematics). In any case, what researchers in the field agree on is the negative economic and social consequences of this

gap, and its associated negative effect on occupational choices and subsequent lower wages for women (Ma, 2008).

From a global standpoint, researchers exploring the gender gap using PISA and TIMSS, have highlighted the comparative sizable gender gap in the Chilean school system. Focusing on TIMSS 2003, Else-Quest et al. (2010) reported that from the 46 nations analysed, Chile, Ghana and Morocco showed a similar pattern in terms of the biggest gender gaps in mathematics achievement in favour of boys. In spite of technical and other limitations of PISA (Goldstein, 2004; Volante, 2016), it has consistently reported a significant gender difference in favour of boys in math (OECD, 2004, 2015) and science (OECD, 2007), and only a small gender difference in favour of girls in reading (OECD, 2007, 2015; OECD/UNESCO-UIS, 2003) among Chilean 15 years-old students. Although Chile experienced one of the largest improvements (12 PISA points) in mathematics from 2006 to 2012, the magnitude of the gender gap remained large, with girls achieving 25 score points less than boys, which is significantly higher than the OECD average of 11 points. Though this gap narrowed in PISA 2015, boys still outperformed girls by 18 score points, which is significantly higher than the OECD average of 8 score points (OECD, 2015). It also remains notable that, after comparing PISA performance of Chile in relation to neighbour countries such as Argentina and Brazil, local researchers have also recognised that “Chile shows the largest differences in favour of males in math and science, and the smallest differences in favour of women in reading, in a very atypical pattern that is clearly disadvantageous to female students” (Manzi, Strasser, San Martín, & Contreras, 2008, pp. 130-131).

From a global standpoint there is pressing priority of reducing the gender gap in mathematics in Chile. This gap has also been recognised by local researchers as an indicator of where

performance is lagging behind. Much of the evidence comes from absolute gender gaps in performance. Given that previous research has stressed that “the bulk past 50 years suggests that the gender gap in mathematics does not exist before children enter school, but is large and significant in the middle school years and beyond” (Fryer & Levitt, 2010, p. 211), it is tempting to attribute gender gaps to schools. While this is true, looking at aggregate gender gaps “ignore[s] the variation in gender differences across schools, thus overlooking the impact of school experiences on gender differences” (Ma, 2008, p. 443). Thus it is necessary to explore the role of schools in explaining differential attainment and progress of boys and girls.

### **Literature review**

Differential effectiveness (DE) has been defined as one of the three sub-themes of School Effectiveness Research (SER)’s consistency, along with stability over time and promotion of different educational outcomes (Sammons, Thomas & Mortimore, 1997). Because schools tend to be inconsistently effective, in order to avoid misleading inferences about the overall effectiveness of schools, DE provides multiple measures according to different types of pupil. The effectiveness of each school varies as a function of the characteristics of their pupils (Leckie, 2008). How effective a school appears to be depends to an extent on whether individual pupil characteristics such as gender, social class and ethnicity are taken into account. DE has recently been described as “some schools may be more effective in promoting the progress of low SES [Socioeconomic Status] than high SES pupils, or boys vs girls, or some ethnic groups more than others. This is termed differential school effectiveness” (Strand, 2016, p. 108). Hence DE refers to internal variation of the school in promoting progress of different groups of students, after taking into account prior attainment and average differences between these groups. DE has also been defined as "the existence of

systematic differences in attainment within schools for different pupil groups (those with different levels of prior attainment or different background characteristics), once the average differences between these groups have been accounted for" (Thomas, Sammons, Mortimore, & Smees, 1997, p. 453). DE is also understood as one of the principles derived from the application of hierarchical linear models to the analysis of student and school data: "schools can have varying effects on different types of students. Some schools may be particularly effective for high-ability students, but not for low-ability students and vice-versa" (Willms & Raudenbush, 1989, p. 214). For these reasons, there is a strong consensus across SER researchers on the need to explore DE within each school as well as overall (Nuttall, Goldstein, Prosser, & Rasbash, 1989) in the efforts to provide more equal educational provisions (Scheerens, 2000).

### ***International Evidence of School Differential Effects***

During the last two decades researchers in the field have pointed out the necessity of exploring how consistent are the school effects in School Effectiveness Research (SER) (Kyriakides, 2004; Leckie, 2008; Sammons, Nuttall, & Cuttance, 1993). The international evidence of school differential effects is reported to be inconclusive, and less stable than evidence concerning school overall effectiveness and progress (Sammons et al., 1993). Conversely, this evidence has been described as "limited and conflicting" (Gray, Peng, Steward, & Thomas 2004, p. 535), and "mixed" (Strand, 2016, p. 109), particularly the evidence referring to gender and ethnicity (Kyriakides, 2004). These conflicting results can be partly explained by the studies' heterogeneity and the inclusion/exclusion of contextual factors. Perhaps this limited evidence is also related to the fact that interpreting differential effects is complex and reflects the difficulty in separating and identifying all the possible

factors that might have an impact on pupils' attainment and progress (Thomas et al., 1997).

Table 1: Gender Differential Effects (DE) value-added studies

| Study                 | Level               | Results  |
|-----------------------|---------------------|--|
| Sammons et al. (1993) | Primary education   | No statistically significant DE for pupils classified by gender, SES or ethnicity  |
| Kyriakides (2004)     | Primary education   | No statistically significant DE for pupils classified by gender or SES in Mathematics  |
| Strand (2010)         | Primary education   | Statistically significant DE for pupils classified by gender   |
| Strand (2016)         | Primary education   | No statistically significant DE for pupils classified by gender in English, Mathematics or Science   |
| Thomas et al. (1997)  | Secondary education | Statistically significant DE for groups of students classified by prior attainment and ethnicity, but not by gender or eligibility for free school meals |
| Gray et al. (2004)    | Secondary education | No statistically significant DE for pupils classified by gender  |

Table 1 summarises the literature review on gender DE according to authors and level of the educational system

Focusing on value-added progress in primary schools, Mortimore, Sammons, Stoll, Lewis and Ecob (1988) analysed the Inner London Education Authority's (ILEA) Junior School Project (JSP) at year 3 and at year 5 and reported differential effects by gender in the progress in Reading in 8 of the 50 (16%) schools analysed. Sammons et al. (1993), re-analysed the JSP and reported differential effectiveness for prior attainment in reading and mathematics, but no statistically significant differential effectiveness for pupils classified by gender, SES or ethnicity. Kyriakides (2004) found no differential effects by pupils classified for gender or social class in mathematics in a sample of primary schools in Cyprus. Strand (2010) reported significant gender DE when analysing a sample of 534,724 pupils from age 7 to 11 nested within 14,289 English primary schools. However, because he found a statistically significant high correlation of 0.98 value-added progress residuals between girls and boys, he concluded that this difference was not substantial. Later on, when elaborating on these results, Strand (2014) asserted that 'there was no evidence of differential school effects in relation to pupils' entitlement to FSM, gender or ethnicity' (p. 292). Two years later he re-examined and

expanded differential effects for different group of pupils (by gender, ethnicity and poverty), for different curriculum subjects (English, mathematics and science) and over time (different cohorts) in a cohort of over 6,000 pupils age 7 and 11 attending 57 primary schools in an inner London borough. He concluded that, although differences between schools in average pupil progress were sizable, there was no evidence of statistically significant gender differential school effectiveness. “More effective schools ‘raised the bar’ but did not ‘close the gap’ suggesting that differences between schools in ‘quality’ play little role in equity gaps” (Strand, 2016, p. 107).

Concentrating on value-added progress in secondary schools, Thomas et al. (1997) reported DE for groups of students classified by prior attainment and ethnicity, but not by gender or eligibility for free school meals. Although Gray et al. (2004) did not find gender DE in most of the secondary schools analysed, they found that schools that were adding value (i.e. progressing above expectations) tended to have differential effects in favour of girls. In short, what previous research has been able to demonstrate is that schools tend to produce DE by prior achievement (Crawford, Macmillan, & Vignoles, 2017; Goldstein, Rasbash, Yang, Woodhouse, Pan, Nuttall, & Thomas, 1993; Sammons et al., 1993; Thomas & Mortimore, 1996; Thomas et al., 1997), but less so by gender (Mortimore et al., 1988; Strand, 2010) or ethnicity (Thomas et al., 1997).

### ***Local Evidence of School Differential Effects***

Evidence of effectiveness may not be generalizable to different settings or new populations (Murillo, 2007; Scheerens, 2015). As the universal validity of SER cannot be taken for granted (Murillo, 2007; Murillo & Román, 2011), it is necessary to explore the role that schools play in the achievement gaps in Latin-America in general and Chile in particular.





At the regional level, Ma (2008) carried out a multilevel analysis of PISA 2000 results.

Although he reported that Chile was one of the few countries that showed variation in gender differences across schools in mathematics performance at one point in time<sup>1</sup>, he did not find statistically significant gender school level variation (differential effects) in absolute levels of attainment. Thus he concluded that the disadvantaged position of Chilean girls could be explained by gender-differentiated processes impacting schooling such as “cultural values, social stereotypes, differential treatments, and culturally specific motivational patterns” (p. 455).

Even though there have been recent value-added studies of student progress in the Chilean educational system focusing on different aspects, such as parental choice (Carrasco & San Martín, 2011), school accountability (San Martín & Carrasco, 2012), the hierarchical structure of the Chilean educational system (Muñoz-Chereau & Thomas, 2016; Troncoso, Pampaka & Olsen, 2016) and recently, the relevance of exploring quantile value added (Page, San Martín, Orellana & González, 2017), these studies have not addressed specifically the gender gap, nor have they explored differential effects in mathematics. Despite their differences, what all previous Chilean value-added studies agree on is the need to take into account the large influence of the schools’ social context. This is also in line with the latest PISA report that stressed that among the participating countries, Chilean differences in raw performance attributable to students’ socio-economic status remain large (23%), in comparison with the average of 15% in OECD countries. Socioeconomic factors systematically explain one of the largest proportions of between-school variance in Chile (OECD, 2015).

---

1  
Ma described a sizable male advantage of 15.99 mean score points, and a standard deviation above 20 score points.

Ma (2008) also proposes that within-school gender differences should be estimated with statistical controls over likely confounding variables. In other words, there is a need to take into account the intersectionality of gender, by considering other factors –such as family background and school SES– when exploring the role of schools in explaining Chilean gender gap in mathematics. Therefore exploring DE in terms of gender in the Chilean context is an innovation, as school effectiveness has been less explored as a relative term –i.e. depending on student grouping– than as an absolute one.

## **Methodology**

### ***Data and sample***

This paper analyses a longitudinal dataset of the Chilean Educational Quality Measurement System (SIMCE) standardized test in mathematics. The SIMCE datasets are cross-sectional, as pupils' academic examinations are not officially tracked over time. This imposes major difficulties, because it is up to researchers to explore the longitudinal nature of the Chilean educational data. Also SIMCE tests are not vertically equated, so interpretation of progress is not straightforward. We followed the cohort of pupils who took the SIMCE test in Year 4 up to Year 8, which in Chile represents the final year of primary education. The dataset is based on merged data from three sources. The first one is the SIMCE tests scores. We use the 4th grade (2005) and 8th grade (2009) results. This dataset was merged with a survey of parents of the students who took the SIMCE tests by researchers at Pontificia Universidad Católica de Chile, who shared the dataset for further analysis (for details of their analysis see San Martín & Carrasco, 2012). The third source of data is administrative records from the Ministry of Education, which provide school-level variables, such as school location and SES. Different from Carrasco and San Martín (2012), the sample analysed in this study

includes private schools and explores the role that municipalities play when estimating Chilean primary schools effectiveness.

In 2005 SIMCE mathematics was applied to 252,481 pupils in Year 4 enrolled in 7,517 schools. In 2009, 239,745 pupils enrolled in 5,814 schools in Year 8 took the SIMCE tests (92.5% of the population). 227,993 of these students took the SIMCE mathematics test in both moments (San Martín & Carrasco, 2012). As the dataset analysed in this study is comprised of 163,044 pupils attending 3,355 primary schools located within 310 municipalities, it represents 90.4% of the municipalities, 57.7% of the schools and 71.5% of the students that took the SIMCE tests in 2009.

### ***Variables***

Tables 2, 3 and 4 present descriptive statistics for the pupil background characteristics (gender and number of books at home), the school background characteristics (School SES), and the normalized SIMCE scores (MathY4 and MathY8) for the whole sample as well as for boys and girls separately. The distributions of normalized MathY4 and MathY8 are fairly symmetric and normally distributed, as can be seen in Figures 1 and 2.

[Figure 1]

[Figure 2]

### ***Models***

The following analysis focuses on differential effects by gender using different 3-level models. The paper compares how the regression coefficients and 2009 Year 8 school effects change across seven models. In order to determine the unique influence of each variable on mathematics pupil achievement and progress age 8–12, and to capture the hierarchical nature

of the data, multi-level regression models were fitted. Analyses were carried out using the MLWin software (V2.36) with pupils' 2009 mathematics age 12 test scores (level 1) grouped within schools (level 2) within municipalities (level 3).

### *Fixed effects*

To model Math 2009, seven multilevel models are compared (See the appendix). In all models only the intercept term (cons) is random at each level:

- (1) Model 0: A null model with an intercept term random at each level. This is the 'null' model and is considered a baseline against which subsequent models are compared.
- (2) Model 1: A Contextualized Attainment Model (CAM) that adds gender as a covariate.
- (3) Model 2: A Contextualized Attainment Model (CAM) that adds books at home as a covariate.
- (4) Model 3: A Contextualized Attainment Model (CAM) that adds School SES as a covariate.
- (5) Model 4: A Contextualized Attainment Model (CAM) that adds gender, Books at home and School SES covariates.
- (6) Model 5: Progress model, controlling for prior attainment in math at age 8 (Year 4). This predictor was centred in the grand mean.
- (7) Model 6: A value added (VA) model including prior attainment in math at age 8 (Year 4) and the student background/characteristics (gender and number of books at home) as well as the school background (School SES) included in previous models.

Table 2: Descriptive statistics for girls and boys

| Variable                | Value      | Frequency | Percent | Minimum | Maximum |
|-------------------------|------------|-----------|---------|---------|---------|
| Gender                  | Girl       | 84,696    | 51.9    | N/A     | N/A     |
|                         | Boy        | 78,345    | 48.1    |         |         |
| Number of books at home | 0 to 50    | 39,873    | 24.5    | N/A     | N/A     |
|                         | 51 to 100  | 98,825    | 60.6    | N/A     | N/A     |
|                         | >100       | 24,065    | 14.8    | N/A     | N/A     |
| School SES              | Low        | 9,014     | 5.5     | N/A     | N/A     |
|                         | Middle-Low | 47,171    | 28.9    | N/A     | N/A     |

|                    |             |         |      |          |         |
|--------------------|-------------|---------|------|----------|---------|
|                    | Middle      | 58,376  | 35.8 | N/A      | N/A     |
|                    | Middle-High | 33,988  | 20.8 | N/A      | N/A     |
|                    | High        | 14,495  | 8.9  | N/A      | N/A     |
| Math_Y8 (Z scores) | N/A         | 163,044 | N/A  | -2.55471 | 2.62680 |
| Math_Y4(Z scores)  | N/A         | 163,044 | N/A  | -3.27767 | 2.00401 |

Table 3: Descriptive statistics for girls

| Variable                | Value       | Frequency | Percent | Mean      | Minimum  | Maximum |
|-------------------------|-------------|-----------|---------|-----------|----------|---------|
| Number of books at home | 0 to 50     | 20,393    | 24.1    | N/A       | N/A      | N/A     |
|                         | 51 to 100   | 51,551    | 60.9    | N/A       | N/A      | N/A     |
|                         | >100        | 12,613    | 14.9    | N/A       | N/A      | N/A     |
|                         | Total       | 84,557    | 99.8    | N/A       | N/A      | N/A     |
| School SES              | Low         | 4,726     | 5.6     | N/A       | N/A      | N/A     |
|                         | Middle-Low  | 24,155    | 28.5    | N/A       | N/A      | N/A     |
|                         | Middle      | 30,927    | 36.5    | N/A       | N/A      | N/A     |
|                         | Middle-High | 17,624    | 20.8    | N/A       | N/A      | N/A     |
|                         | High        | 7,264     | 8.6     | N/A       | N/A      | N/A     |
| Math_Y8 (Z scores)      | N/A         | 84,696    | N/A     | -.0999964 | -2.54156 | 2.62680 |
| Math_Y4(Z scores)       | N/A         | 84,696    | N/A     | -.0787908 | -3.23643 | 2.00401 |

Table 4: Descriptive statistics for boys

| Variable                | Value       | Frequency | Percent | Mean     | Minimum  | Maximum |
|-------------------------|-------------|-----------|---------|----------|----------|---------|
| Number of books at home | 0 to 50     | 19,479    | 24.9    | N/A      | N/A      | N/A     |
|                         | 51 to 100   | 47,272    | 60.3    | N/A      | N/A      | N/A     |
|                         | >100        | 11,452    | 14.6    | N/A      | N/A      | N/A     |
|                         | Total       | 78,203    | 99.8    | N/A      | N/A      | N/A     |
| School SES              | Low         | 4,288     | 5.5     | N/A      | N/A      | N/A     |
|                         | Middle-Low  | 23,015    | 29.4    | N/A      | N/A      | N/A     |
|                         | Middle      | 27,448    | 35.0    | N/A      | N/A      | N/A     |
|                         | Middle-High | 16,363    | 20.9    | N/A      | N/A      | N/A     |
|                         | High        | 7,231     | 9.2     | N/A      | N/A      | N/A     |
| Math_Y8 (Z scores)      | N/A         | 78,345    | N/A     | .1081128 | -2.55471 | 2.62680 |
| Math_Y4(Z scores)       | N/A         | 78,345    | N/A     | .0851767 | -3.27767 | 2.00401 |

### ***Random effects***

Focusing on the random part of the models, the variation at each hierarchical level (pupils, schools and municipalities) is explored.

### **Results**

In order to describe the gender gap in mathematics Y4 and Y8, the means of Math\_Y4(Z scores) and Math\_Y8(Z scores) were compared. The mean for boys in Math Y4 and Math Y8 were .085 SD units and .108 SD units, respectively. The mean for girls in Math Y4 and Math

Y8 were  $-.079$  SD units and  $-.099$  SD units, respectively. In other words, the gender gap in favor of boys in mathematics performance increased from 0.77 in Year 4, to 0.89 in Year 8.

Table 5: Fixed part

|  | Model 0<br>Empty/raw<br>3-Level<br>model<br>variance<br>component<br>model | Model 1<br>Model 0 +<br>gender | Model 2<br>Model 0 +<br>books at<br>home | Model 3<br>Model 0 +<br>School SES | Model 4<br>Model 0 +<br>gender +<br>books at<br>home+<br>School SES | Model 5<br>Model 0 +<br>prior<br>attainment | Model 6<br>Model<br>0+gender +<br>books at<br>home +<br>School SES<br>+ prior<br>attainment | Model 6a                       |
|--|--|--------------------------------|--|------------------------------------|---|---|---|--------------------------------|
| Variable<br>Fixed part<br>Cons Intercept | -0.163 [0.02]  | -0.051 [0.02]                  | -0.275 [0.02]                            | -0.571 [0.02]                      | -0.531 [0.02]   | -0.062 [0.01]                               | -0.185 [0.02]   | -0.186 [0.02]                  |
| MathsY4(Z score)                         | N/A  | N/A                            | N/A                                      | N/A                                | N/A   | 0.612 [0.00]                                | 0.598 [0.00]  | 0.598 [0.00]                   |
| Girl                                     | N/A  | -0.214 [0.00]                  | N/A                                      | N/A                                | -0.214 [0.00]   | N/A   | -0.118 [0.00]   | -0.186 [0.00]<br>-0.117 [0.00] |
| 51-100 books at home                     | N/A  | N/A                            | 0.148 [0.00]                             | N/A                                | 0.141 [0.00]  | N/A   | 0.059 [0.00]  | 0.059 [0.00]                   |
| >100 books at home                       | N/A  | N/A                            | 0.244 [0.01]                             | N/A                                | 0.231 [0.01]  | N/A   | 0.103 [0.00]  | 0.103 [0.00]                   |
| Middle Low SES                           | N/A  | N/A                            | N/A                                      | 0.112 [0.02]                       | 0.089 [0.02]  | N/A   | -0.01 [0.02]  | -0.01 [0.02]                   |
| Middle SES                               | N/A  | N/A                            | N/A                                      | 0.513 [0.02]                       | 0.467 [0.02]  | N/A   | 0.150 [0.02]  | 0.151 [0.02]                   |
| Middle-High SES                          | N/A  | N/A                            | N/A                                      | 1.026 [0.03]                       | 0.964 [0.03]  | N/A   | 0.398 [0.02]  | 0.399 [0.02]                   |
| High SES                                 | N/A  | N/A                            | N/A                                      | 1.592 [0.03]                       | 1.503 [0.03]  | N/A   | 0.727 [0.02]  | 0.735 [0.03]                   |

Figure 1

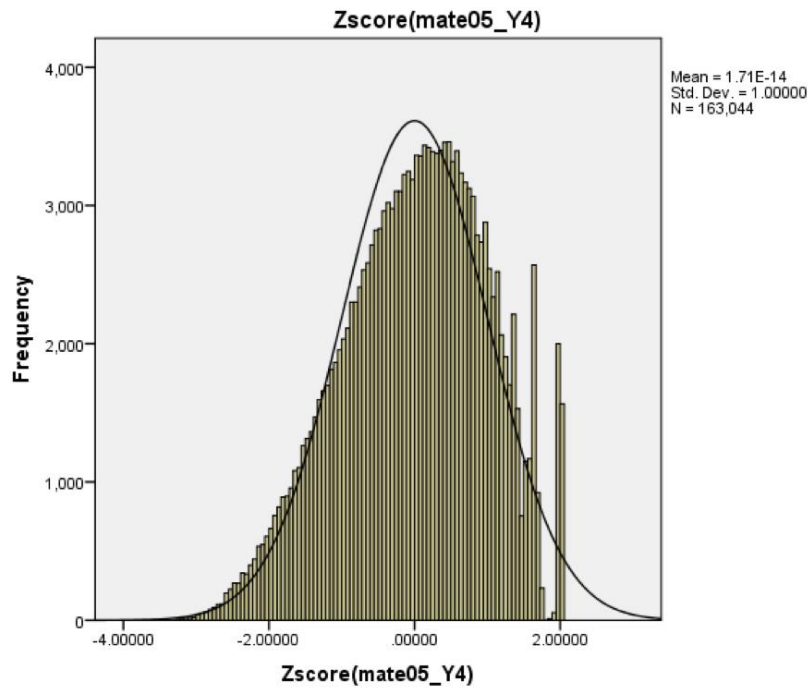
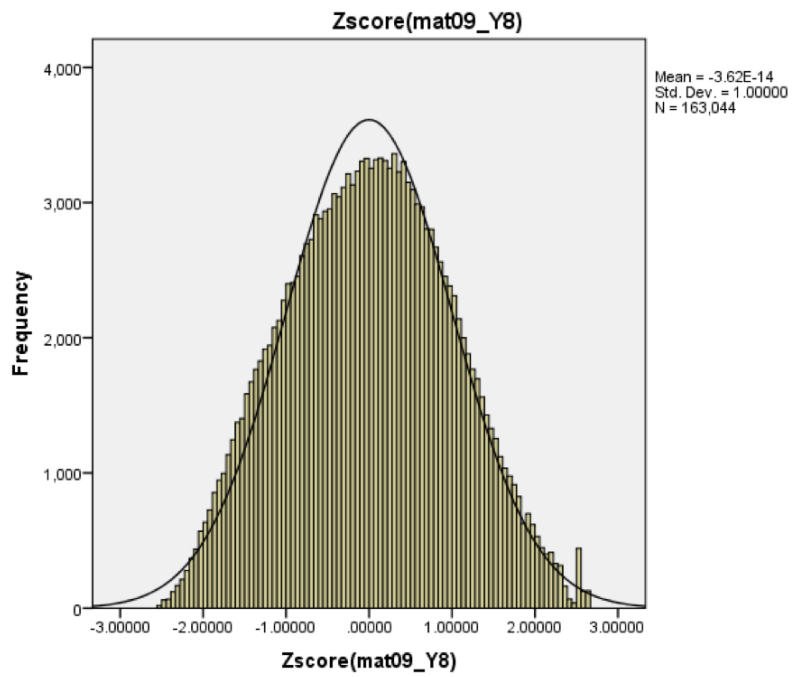


Figure 2



*Pupil Background Characteristics*

Focusing on Table 5, it is clear that the effect for gender was in the expected direction. Girls achieved significantly less than boys in mathematics Year 8 ( $-0.21$  SD units) after adjusting for school SES and number of books at home (Model 4). Girls also made significantly less progress than boys ( $-0.12$  SD units) after adjusting for prior achievement, school socioeconomic status (SES) and number of books at home (Model 6). Similarly, the effect of number of books at home was in the expected direction. Model 4 shows how pupils with 51-100 books, and with more than 100 books at home achieved significantly more in math at age 14 than those with 0 to 50 books at home ( $0.14$  and  $0.23$  SD units, respectively). Model 6 shows that pupils with more books at home also progressed significantly more ( $0.06$  and  $0.10$  SD units) in math at age 14 than those with 0 to 50 books at home, even after accounting for other covariates. This is in line with previous research that has found that the presence of books at home is positively correlated with the socioeconomic status of the family and is considered a strong predictor of pupils' achievement (Sullivan & Brown, 2013).

#### *School Background Characteristics*

Although the effect of School SES on math achievement is as expected, with pupils achieving up to 1.5 SD units higher scores in High SES schools in comparison with Low SES schools, even after accounting for other covariates (Model 4), the progress of Middle-Low SES schools is lower but not statistically different from Low SES schools (Model 6). This result, at first glance counterintuitive, is in line with Valenzuela, Bellei and Allende (2016), who found that, over the last decade, school trajectories were more likely to improve among primary Chilean schools with lower SES.



*Random effects*

Table 6 presents the random effects

Table 6: Random part

|                                  | Model 0<br>Empty/raw<br>3-Level model<br>variance<br>component<br>model | Model 1<br>Model 0 +<br>gender | Model 2<br>Model 0 +<br>books at home | Model 3<br>Model 0 +<br>School SES | Model 4<br>Model 0 +<br>gender +<br>School SES | Model 5<br>Model 0+ prior<br>attainment | Model 6<br>Model 0+<br>gender + books<br>at home +<br>School SES +<br>prior attainment |
|----------------------------------|---|--------------------------------|---------------------------------------|------------------------------------|--|---|--|
| Between municipalities intercept | 0.047 [0.01]  | 0.047 [0.01]                   | 0.042 [0.01]                          | 0.012 [0.00]                       | 0.012 [0.00]                                   | 0.012 [0.00]                            | 0.005 [0.00]   |
| Between schools intercept        | 0.277 [0.007]   | 0.279 [0.007]                  | 0.258 [0.007]                         | 0.107 [0.003]                      | 0.106 [0.003]                                  | 0.099 [0.003]                           | 0.061 [0.002]  |
| Between pupils Intercept         | 0.646 [0.002]   | 0.635 [0.00]                   | 0.641 [0.002]                         | 0.646 [0.002]                      | 0.631 [0.002]                                  | 0.369 [0.001]                           | 0.365 [0.001]  |
| Total variance                   | 0.97  | 0.961                          | 0.941                                 | 0.765                              | 0.749  | 0.48                                    | 0.431  |
| % of variance explained          |   |                                |                                       |                                    |  |   |  |
| Municipality                     | N/A   | 0%                             | 10.63%                                | 74.46%                             | 74.46%   | 74.46%                                  | 89.36%   |
| School                           | N/A   | 0.72%                          | 6.85%                                 | 61.37%                             | 61.73%   | 64.25%                                  | 77.97%   |
| Pupil                            | N/A   | 1.70%                          | 0.77%                                 | 0%                                 | 2.32%  | 71.13%                                  | 43.49%   |
| Total                            | N/A   | 0.92%                          | 2.98%                                 | 21.13%                             | 22.78%   | 50.51%                                  | 55.56%   |
| % of variance attributable to    |   |                                |                                       |                                    |  |   |  |
| Municipality                     | 4.84%   | 4.89%                          | 4.46%                                 | 1.56%                              | 1.6%   | 2.5%                                    | 1.16%  |
| School                           | 28.56%  | 29.03%                         | 27.41%                                | 13.98%                             | 14.15%   | 20.62%                                  | 14.15%   |
| Pupil                            | 66.59%  | 66.1%                          | 68.12%                                | 84.44%                             | 84.24%   | 76.87%                                  | 84.68%   |
| 2-log likelihood                 | 401448.251  | 398899.599                     | 399441.759                            | 398501.540                         | 394065.867                                     | 308680.799                              | 304883.452   |
| Sample size                      |   |                                |                                       |                                    |  |   |  |
| Municipality                     | 310   | 310                            | 310                                   | 310                                | 310  | 310                                     | 310  |
| School                           | 3355  | 3355                           | 3355                                  | 3355                               | 3355   | 3355                                    | 3355   |
| Pupil                            | 163044  | 163041                         | 162763                                | 163044                             | 162760   | 163044                                  | 162760   |

There are five key points to highlight from Table 6. Firstly, the estimated size of the municipality level is smaller than the school effects, but nevertheless of substantive relevance. The variance attributable to municipality ranged from 5% (Raw) to 1.2% (VA). This in line with previous research that has stated that “any fair effort to assess the quality of schools in Chile should not neglect the municipality level because the regional context (such as socio-economic or geographical factors) is also significantly affecting the possibilities of a school of being more or less effective” (Muñoz-Chereau & Thomas, 2016, pp. 42-43).

Secondly, the proportion of variance at the school level in the null model is around 28.5%. Whilst little proportion of this variation is explained by the student background/characteristics (0.7% by gender and 6.8% by books at home), a substantial proportion of this is explained by school background characteristics and prior attainment at age 8 (Year 4). This is highlighted by the increase in the explained school level variance of 61% in Model 3, and 64% in the Progress model 5. The full VA model (Model 6) explains an even greater proportion, accounting for around 78% of the school level variance. This substantial increase in the variance

reduction for the prior attainment only model (Model 5) against the VA model (Model 6), highlights the relevance of considering school background variables and pupil background characteristics when exploring variation in pupil progress between Chilean primary schools. Thirdly, after accounting for pupil background characteristics, school background and pupil prior attainment (VA Model 6), there is still 14% of variability between schools or school effects on pupil progress age 8–12. Fourthly, the magnitude of the primary school effect seems higher than what previous research has reported for industrialised countries. The proportion of variance attributable to the school was reduced from 28.5% (Model 0: Null Model) to 14% (Model 6: VA Model). However, this is in line with previous studies that have shown that school effects are bigger over longer periods of time, conducted over 5 to 10 years (Leckie, 2008; Thomas et al., 1997; Thomas, Peng, & Gray, 2007; Willms & Raudenbush, 1989). It is also in line with Bosker and Witziers (1996, in Opdenakker & Van Damme, 2000), who found in a meta-analysis of 103 MLM (Multilevel Modelling) studies that Raw score variation was bigger in Third World countries in comparison with industrialised ones. It is also important to bear in mind that most of the SER studies have been conducted exclusively in public schools, whilst this dataset includes public, mixed and private schools. Finally, after comparing the log-likelihood for Model 0, Model 4, Model 5 and Model 6, the reduction of deviance of IGLS (Iterative Generalised Least Squares) suggests that Model 6 is significantly better than previous models. More precisely, from Model 0 to Model 1, it improved by 2548.652 points; Model 1, Model 2 and Model 3 are complementary, as they are modelling the effect of gender, books at home and School SES separately. From Model 0 to Model 4, a Contextualised Attainment Model that jointly controlled for these variables, IGLS Deviance reduced by 7382.384; from Model 4 to Model 5 it reduced by 4435.673; and from Model 5 to Model 6 the reduction was by 3797.347.

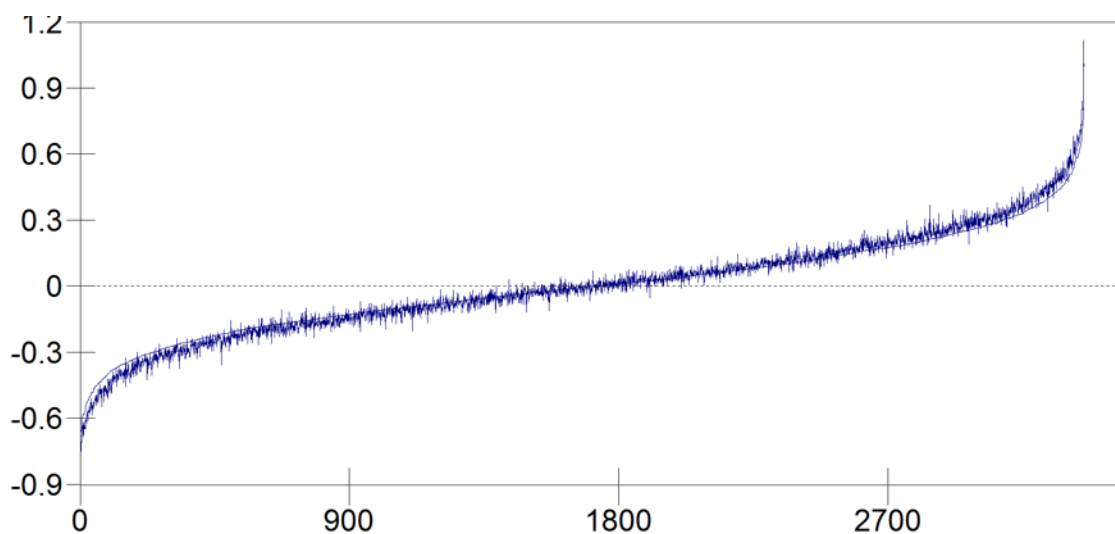
### *Differential school effectiveness by pupil characteristics*

Given the remaining variance in Maths progress attributable to school differences, now the question of whether schools that progress well with boys tend to progress well with girls needs to be addressed.

In order to explore differential effects by gender on value-added progress, Model 6a extends Model 6 by including separate intercept for boys and girls in the random part at the school level. This specification of the random part improved by reducing the log-likelihood by 101.967 points compared with the same model with only random intercepts (Model 6). [Table 7].

From Table 7 it can be stated that there is a significantly small school variation in progress (0.055 for girls and 0.069 for boys) in relation to pupils classified by gender. Moreover, girls' value-added progress within schools is slightly less variable than boys' progress. To visualise this, Figure 3 plots the school residuals caterpillar plots for Model 6a estimates for girls and boys' progress.

Figure 3



In Figure 3 the straight line describes the residuals for girls and the zig-zag line represents the residuals for boys. Interestingly, in the schools that are not adding value (left-side of the plot) girls appear to be doing better than boys, whilst in the schools that are adding value (right-side), the opposite seems true.

Table 8: Classification of the schools value-added progress according to Model 6a residuals for girls and boys

| Statistically significant residuals for girls |          |         |          |          |       |
|---|----------|---------|----------|----------|-------|
| Statistically significant residuals For boys  |          | Average | Negative | Positive | TOTAL |
|   | Average  | 2,067   | 33       | 34       | 2,134 |
|   | Negative | 46      | 554      | 0        | 600   |
|   | Positive | 31      | 0        | 590      | 621   |
|   | TOTAL    | 2,144   | 587      | 624      | 3,355 |

Note: Average: schools for whom their residuals for girls and boys cannot be reliably distinguished from an average school. Negative: schools with residuals statistically significantly below the average school. Positive: schools with residuals statistically significantly above the average school.

Table 8 presents a classification of the statistically significant residuals for girls and boys value-added progress according to Model 6a for the 3,355 primary schools, after considering the 95% confidence intervals. School residuals in 61.6% of the schools (2,067) cannot be reliably distinguished from the progress made by girls and boys on an average school. Whilst in 590 schools (17.6%), boys and girls progressed above expectations, in 554 schools (16.5%) both progressed significantly below expectations. In the remaining 144 of the schools (4.2%), there is evidence of a significant differential effect in the progress made by girls or boys. In other words, in 144 Chilean primary schools the progress made by boys is at the expense of girls, or vice versa: in 46 schools boys progressed below expectations whilst girls progressed as expected; in 31 schools boys progressed above expectations whilst girls progresses as expected; in 33 schools girls progressed below expectations whilst boys

progressed as expected, and in the remaining 34 schools, girls progressed above expectations whilst boys progressed as expected. Given that the correlation of the estimates of school effects on progress for girls and boys is almost perfect ( $r= 0.99$ ), it is possible to conclude that the evidence of school gender differential school effect in Chilean primary schools is not substantial.

## **Discussion**

This study explored the role that Chilean primary schools play in explaining the achievement gap in mathematics. The growing awareness that the universal validity of SER across different country contexts cannot be taken for granted is supported by our results, as they extend SER knowledge base in several ways. First of all, the study found extremely small gender differential effects in student progress in mathematics in the Chilean educational system. This finding is in line with all of the previous studies that have explored gender differential effects in primary schools (Goldstein et al., 1993; Kyriakides, 2004; Sammons et al., 1993; Strand, 2010, 2014, 2016), What appears relevant to consider, however, is that gender differentials can be better understood as context specific. For example, when considering the gender differential effect reported in a minority of secondary schools, Gray et al. (2004) found that schools that were adding value (i.e. progressing above expectations) tended to have differential differences in favour of girls. The opposite was found in this study: in Chilean primary schools that were low performing or not adding value, girls appeared to be doing better than boys, whilst in the schools that were better performing or adding value, the opposite was true. In other words, differences in schools in terms of effectiveness, appear to be playing a significant role in the gender gap. Crucially, girls' disadvantaged position in performance and progress in math is located at the top of the distribution. As this study does not provide information about why this might be the case,

future qualitative research could provide timely information to enhance teacher training programs and local policies oriented to promote gender equality in the Chilean educational system.

Also, the statistically significant high correlation of 0.98 between girls and boys residuals in a sample of primary schools in England lead Strand (2010, 2014) to conclude that there was not a substantial gender differential effect. Given that this correlation is similar in magnitude to the correlation of 0.99 between girls and boys value-added progress residuals found in Chilean primary schools, this study also concludes that Chilean primary schools tend not to produce gender differential effects. Moreover, it is important to note that the great majority of Chilean primary schools maintain (instead of reduce or improve) the achievement gap between girls and boys when comparing progress measures. Additionally, most of the mathematics achievement variation lies between students within schools. In line with Strand (2016), I conclude that although Chilean primary schools have a role to play in reducing achievement gaps between girls and boys, policymakers should mainly address this issue by looking at causes beyond the schools. This is also in line with previous studies that have explored the Chilean gender gap in favour of boys in mathematics using PISA (Ma, 2008; Manzi et al., 2008; OECD, 2007, 2015; OECD/UNESCO-UIS, 2003) and TIMSS (Else-Quest et al., 2010). These studies have highlighted the negative influences of cultural values, differential treatments, culturally specific motivational patterns and gender stereotypes regarding math perpetuated by the media, leading girls to be less confident irrespective of their own math ability. In this direction, any successful policy oriented to improve gender educational equality will need to counteract this negative effect by, for example, highlighting the achievements of relevant role models in Science, Technology, Engineering and Mathematics (STEM), such as Maria Teresa Ruiz, the first female Chilean professor that

obtained a 2017 L'Oréal-UNESCO For Women in Science Awards, as well as supporting local practitioners, such as Colectivo de mujeres matemáticas [Collective of Women Mathematicians], a group formed by female Chilean mathematicians in 2014, oriented to tackle the gender gap and the differences of opportunities faced by women by delivering talks in secondary schools. Initiatives such as these are significant steps towards making maths accessible and relevant for Chilean girls.

Finally, though this study proposes a novel approach in its exploration of gender differential effects in the Chilean context, there are a number of limitations that need to be highlighted. The first is the lack of causality in non-experimental settings. The analysis conducted in this research aimed to provide empirical descriptions that are exploratory, highlighting more tendencies than certainties. Secondly, inferences from this sample cannot be made to a larger population. Thirdly, despite research showing that the gender gap in favour of boys remains considerable in terms of attitudes and affect (Frenzel, Pekrun, & Goetz, 2007), given the lack of data on non-academic variables in the SIMCE dataset available, this study is limited to pupils' cognitive performance in maths. It is worth exploring in the future the impact of schools in terms of differential effects by subjects, cohorts and pupils classified by age, ethnicity and SES, which might be significant in the Chilean educational system. Future research also need to test the potential influence of school effects over time, the effect of students' mobility between schools, and more variables to account for pupil SES.

## **References**

Aikman, S., & Unterhalter, E. (2013). Gender equality, capabilities and the terrain of quality education. In L. Tikly and A.M. Barrett (Eds.), *Education Quality and Social Justice in the Global South: Challenges for Policy, Practice and Research* (pp. 25-39). London: Routledge.



Carrasco, A., & San Martín, E. (2011). Are quasi-markets in education meeting their policy purposes in Chile? Re-examining empirical hypothesis from value-added models. *Informes Técnicos MIDE UC*. Retrieved from <http://mideuc.cl/wp-content/uploads/2011/09/it1104.pdf>

Crawford, C., Macmillan, L., & Vignoles, A. (2017). When and why do initially high-achieving poor children fall behind?. *Oxford Review of Education*, 43, 88-108.  
doi:10.1080/03054985.2016.1240672

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin, American Psychological Association*, 136, 103–127. Doi:10.1037/a0018053

Frenzel, A.C., Pekrun, R., & Goetz, T. (2007) Girls and mathematics —A “hopeless” issue? A control-value approach to gender differences in emotions towards mathematics. *Eur J Psychol Educ*. 22: 497. doi: 10.1007/BF03173468

Fryer, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-240. Retrieved from <https://scholar.harvard.edu/fryer/publications/empirical-analysis-gender-gap-mathematics>

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford review of*

*education*, 19(4), 425-433. Retrieved from

<http://www.tandfonline.com/doi/abs/10.1080/0305498930190401>

Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: principles, policy and practice*, 11(3), 319-330. Retrieved from <http://dx.doi.org/10.1080/0969594042000304618>

González de San Román, A., & De La Rica, S. (2012). *Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes*. Discussion Papers Series. IZA DP No. 6338. Retrieved from <http://ftp.iza.org/dp6338.pdf>

Gray, J., Peng, W. J., Steward, S., & Thomas, S. (2004). Towards a typology of gender-related school effects: some new perspectives on a familiar problem. *Oxford Review of Education*, 30(4), 529-550. Retrieved from <http://dx.doi.org/10.1080/0305498042000303991>

Kyriakides, L. (2004). Differential school effectiveness in relation to sex and social class: Some implications for policy evaluation. *Educational Research and Evaluation*, 10(2), 141-161. Retrieved from <http://www.tandfonline.com/doi/abs/10.1076/edre.10.2.141.27907>

Leckie, G. (2008) Multilevel Modelling of school differences in educational achievement. (Unpublished doctoral dissertation). University of Bristol, Bristol.

Ma, X. (2008). Within-school gender gaps in reading, mathematics, and science literacy. *Comparative Education Review*, 52(3), 437-460. doi: 10.1086/588762

Manzi, J.; Strasser, K.; San Martín, E., & Contreras, D. (2008). Quality of Education in Chile.

Manuscript, MIDE-UC. Retrieved from

[https://www.researchgate.net/publication/265405330\\_Quality\\_of\\_Education\\_in\\_Chile](https://www.researchgate.net/publication/265405330_Quality_of_Education_in_Chile)

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988) *School Matters: The junior years*. London: Open Books.

Muñoz-Chereau, B., & Thomas, S. M. (2016). Educational effectiveness in Chilean secondary education: comparing different ‘value added’ approaches to evaluate schools. *Assessment in Education: Principles, Policy & Practice*, 23(1), 26-52.

Murillo, J. (2007). School Effectiveness Research in Latin America. In T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement*. Part I. (pp. 75-92). The Netherlands: Springer.

Murillo, J., & Román, M. (2011). The School or the Cradle? Evidence of their Contribution to Students’ Performance in Latin America. Multilevel Study about the estimation of School Effects. *Profesorado*.15(3), 26-50. Retrieved from

<https://recyt.fecyt.es/index.php/profesorado/article/view/41445>

Nuttall, D., Goldstein, H., Prosser, R., & Rasbash, J. (1989) Differential School Effectiveness. *International journal of educational research*, Special issue Developments in school effectiveness research 13(7),769-776. doi.org/10.1016/0883-0355(89)90027-X

Organization for Economic Co-operation and Development/United Nations

Educational Scientific and Cultural Organization-UNESCO Institute of Statistics (2003).

*Literacy Skills for the World of Tomorrow - Further results from PISA 2000*. Retrieved from <http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33690591.pdf>

Organización para la Cooperación y el Desarrollo Económicos [OCDE]. (2004). Revisión de políticas nacionales de educación Chile [Review of National Policies in Education in Chile]. Paris: OCDE. Retrieved from

[http://www.facso.uchile.cl/psicologia/epe/\\_documentos/GT\\_cultura\\_escolar\\_politica\\_educativa/recursos%20bibliograficos/articulos%20relacionados/oecd\(2004\)revisiondepoliciaeducacionenchile.pdf](http://www.facso.uchile.cl/psicologia/epe/_documentos/GT_cultura_escolar_politica_educativa/recursos%20bibliograficos/articulos%20relacionados/oecd(2004)revisiondepoliciaeducacionenchile.pdf)

Organization for Economic Co-operation and Development (2007) *PISA 2006: Science Competencies for Tomorrow's World* . Vol 1. Retrieved from

<http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/pisa2006-publications.htm>

Organization for Economic Co-operation and Development (2015) *Education at a Glance*.

Retrieved from [http://www.oecd-ilibrary.org/education/education-at-a-glance-](http://www.oecd-ilibrary.org/education/education-at-a-glance-2015/percentage-of-low-achievers-in-mathematics-reading-and-science-by-gender-pisa-2012_eag-2015-table110-en)

[2015/percentage-of-low-achievers-in-mathematics-reading-and-science-by-gender-pisa-2012\\_eag-2015-table110-en](http://www.oecd-ilibrary.org/education/education-at-a-glance-2015/percentage-of-low-achievers-in-mathematics-reading-and-science-by-gender-pisa-2012_eag-2015-table110-en)

Opdenakker, M., & Van Damme, J. (2000): *Effects of Schools, Teaching Staff and Classes on Achievement and Well-Being in Secondary Education: Similarities and Differences*

Between School Outcomes. *School Effectiveness and School Improvement*. 11(2),165-196.

Retrieved from [http://www.tandfonline.com/doi/abs/10.1076/0924-3453\(200006\)11:2;1-](http://www.tandfonline.com/doi/abs/10.1076/0924-3453(200006)11:2;1-)

Q;FT165

Page, G., San Martín, E., Orellana, J., & González, J. (2017). Exploring complete school effectiveness via quantile value added. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 315-340. doi:10.1111/rssa.12195

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential School Effectiveness: Results from a reanalysis of the Inner London Education Authority's Junior School Project data, *British Educational Research Journal*, 19(4), 381- 405. Retrieved from <http://dx.doi.org/10.1080/0141192930190407>

Sammons, P., Thomas, S., & Mortimore, P. (1997) *Forging links: Effective Schools and Effective Departments*. London: Paul Chapman Publishing.

San Martín, E., & Carrasco, A. (2012). Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa [Classification of schools in the new educational institutions: contribution of value-added models for fair accountability] (*Vol. 53*). Centro de Políticas Públicas. Retrieved from <http://politicaspUBLICAS.uc.cl/wp-content/uploads/2015/02/descargar-clasificacion-de-escuelas-en-la-nueva-institucionalidad-educativa.pdf>

Scheerens, J. (2000) Improving School Effectiveness. *Fundamentals of Educational Planning. International Institute of Educational Planning*. N. 68. Paris: UNESCO. Retrieved from <https://pdfs.semanticscholar.org/e206/6a565e5669297be5cc61afe2f3a32c650966.pdf>

Scheerens, J. (2015). Theories on educational effectiveness and ineffectiveness. *School effectiveness and school improvement*, 26(1), 10-31.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement*, 21(3), 289-314. Retrieved from <http://dx.doi.org/10.1080/09243451003732651>

Strand, S. (2014) School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11, *Oxford Review of Education*, 40:2, 223-245, DOI:10.1080/03054985.2014.891980

Strand, S. (2016). Do some schools narrow the gap? Differential school effectiveness revisited. *Review of Education*. 4(2), 107–144. doi: 10.1002/rev3.3054

Sullivan, A., & Brown, M. (2013) Social inequalities in cognitive scores at age 16: The role of reading. CLS Working Paper 2013/10. London: Centre for Longitudinal Studies. Retrieved from [http://discovery.ucl.ac.uk/1473708/1/Sullivan\\_CLS%20WP%202013%20\(10\)%20Social%20inequalities%20in%20cognitive%20scores%20at%20age%2016%20The%20role%20of%20reading%20-%20Alice%20Sullivan%20and%20Matt%20Brown%20Sept%202013.pdf](http://discovery.ucl.ac.uk/1473708/1/Sullivan_CLS%20WP%202013%20(10)%20Social%20inequalities%20in%20cognitive%20scores%20at%20age%2016%20The%20role%20of%20reading%20-%20Alice%20Sullivan%20and%20Matt%20Brown%20Sept%202013.pdf)

Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary-school effectiveness. *Research papers in education*, 11(1), 5-33. Retrieved from <http://dx.doi.org/10.1080/0267152960110103>

Thomas, S. (1998). Value-added measures of school effectiveness in the United Kingdom. *Prospects*, 28(1), 91-108.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997) Differential Secondary School Effectiveness: comparing the performance of different pupil groups. *British Educational Research Journal*, 23:4, 451-469. doi:10.1080/0141192970230405

Thomas, S., Peng, W.J., & Gray, J. (2007). Modelling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, 33(3),261–295. Retrieved from <http://dx.doi.org/10.1080/03054980701366116>

Troncoso, P., Pampaka, M., & Olsen, W. (2016). Beyond traditional school value-added models: a multilevel analysis of complex school effects in Chile, *School Effectiveness and School Improvement*, 27:3, 293-314, DOI: [10.1080/09243453.2015.1084010](https://doi.org/10.1080/09243453.2015.1084010)

Valenzuela, J. P., Bellei, C., & Allende, C. (2016). Measuring systematic long-term trajectories of school effectiveness improvement. *School Effectiveness and School Improvement*, 27(4), 473-491. Retrieved from <http://dx.doi.org/10.1080/09243453.2016.1150861>

Volante, L. (Ed.). (2016). *The intersection of international achievement testing and educational policy: Global perspectives on large-scale reform*. NY: Routledge.

Willms, D., & Raudenbush, S. (1989) A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability. *Journal of Educational Measurement*, 26(3), 209-232.doi: 10.1111/j.1745-3984.1989.tb00329.x